Identifying cardiac surgery operations in hospital episode statistics administrative database, with an OPCS-based classification of procedures, validated against clinical data

Bortolussi, Giacomo; McNulty, David; Waheed, Hina; Mawhinney, Jamie A; Freemantle, Nick; Pagano, Domenico

DOI:
10.1136/bmjopen-2018-023316

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (Harvard):
Bortolussi, G, McNulty, D, Waheed, H, Mawhinney, JA, Freemantle, N & Pagano, D 2019, 'Identifying cardiac surgery operations in hospital episode statistics administrative database, with an OPCS-based classification of procedures, validated against clinical data', *BMJ open*, vol. 9, no. 3, e023316. https://doi.org/10.1136/bmjopen-2018-023316

BMJ Open

Identifying cardiac surgery operations in hospital episode statistics administrative database, with an OPCS-based classification of procedures, validated against clinical data

Giacomo Bortolussi,[1] David McNulty,[1] Hina Waheed,[1] Jamie A Mawhinney,[1] Nick Freemantle,[1,2] Domenico Pagano[1,3]

To cite: Bortolussi G, McNulty D, Waheed H, *et al*. Identifying cardiac surgery operations in hospital episode statistics administrative database, with an OPCS-based classification of procedures, validated against clinical data. *BMJ Open* 2019;**9**:e023316. doi:10.1136/bmjopen-2018-023316

► Prepublication history for this paper is available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2018-023316).

Received 6 April 2018
Revised 20 January 2019
Accepted 31 January 2019

[1]Quality and Outcome Research Unit, University Hospital Birmingham, Birmingham, UK
[2]Institute of Clinical Trials and Methodology, University College London, London, UK
[3]Institute of Cardiovascular Sciences, University of Birmingham, Birmingham, UK

Correspondence to
Domenico Pagano;
domenico.pagano@uhb.nhs.uk

ABSTRACT

Objectives Administrative databases with dedicated coding systems in healthcare systems where providers are funded based on services recorded have been shown to be useful for clinical research, although their reliability is still questioned. We devised a custom classification of procedures and algorithms based on OPCS, enabling us to identify open heart surgeries from the English administrative database, Hospital Episode Statistics, with the objective of comparing the incidence of cardiac procedures in administrative and clinical databases.

Design A comparative study of the incidence of cardiac procedures in administrative and clinical databases.

Setting Data from all National Health Service Trusts in England, performing cardiac surgery.

Participants Patients classified as having cardiac surgery across England between 2004 and 2015, using a combination of procedure codes, age >18 and consultant specialty, where the classification was validated against internal and external benchmarks.

Results We identified a total of 296 426 cardiac surgery procedures, of which majority of the procedures were coronary artery bypass grafting (CABG), aortic valve replacement (AVR), mitral repair and aortic surgery. The matching at local level was 100% for CABG and transplant, >90% for aortic valve and major aortic procedures and >80% for mitral. At national level, results were similar for CABG (IQR 98.6%–104%), AVR (IQR 105%–118%) and mitral valve replacement (IQR 86.2%–111%).

Conclusions We set up a process which can identify cardiac surgeries in England from administrative data. This will lead to the development of a risk model to predict early and late postoperative mortality, useful for risk stratification, risk prediction, benchmarking and real-time monitoring. Once appropriately adjusted, the system can be applied to other specialties, proving especially useful in those areas where clinical databases are not fully established.

Strengths and limitations of this study

- We produced a detailed classification of OPCS codes relevant to cardiac surgery.
- We used the classification to identify cardiac surgery operations performed in England over a long period from an administrative database.
- Compared with previous studies, we achieved a higher level of detail in our classification, with the aim of analysing specific subgroups of operations.
- The comparison with clinical data provided several layers of control over the quality of our work.
- Information in administrative databases largely depends on the quality of the source, the clinical notes.

INTRODUCTION

Administrative data are collected by organisations involved in healthcare for the purposes of registration, transaction and record keeping during the delivery of a service. Importantly, such information is also used to provide reimbursement for hospitals under the payment by results (PbR) system. This is achieved through the translation of information about patients' diagnoses and any interventions they received into codes in a standardised format by formally trained clinical coders. In the UK, OPCS Classification of Interventions and Procedures version 4.7 is the system used to code interventions, and ICD-10 the system for diagnoses. Each admission may contain several episodes, corresponding to the care provided under a specific consultant, and within each episode, the aforementioned codes are recorded.

Employing administrative data for clinical research has been proposed on numerous occasions; however, not being intended as a primary source for such purposes, their use remains controversial. Many of the issues raised by Aylin *et al* regarding the reliability of the data, whether clinical records can be coded accurately and used in clinical research

Table 1 Classification of OPCS codes by group of procedures

Classification	OPCS description
Principal codes: at least one	
CABG	K4[01234567][123489], K45[56], K4[67]5
Valve surgery	K2[5678][1234589], K276, K3[01][123489], K34[12345689], K36[12], K38[123]
Aortic root surgery	K33[134589]
Aortic surgery	L1[89][289], L20[289], L212, L22[1489], L23[15689], L25[124589]
Aortic surgery (ascending)	K38[56], K551, L1[89]1, L2[01]1
Other cardiac surgery (requiring opening of the pericardium)	K12[1234589], K22[129], K23[123689], K249, K38[89], K48[123489], K53[1289], K55[345689], K67[189], K69[1289], K71[123489], L80[189], T031
AF surgery	K22[38], K52[12345689]
ASD closure	K10[1234589]
VSD closure	K11[123456789]
ECMO	X581
LV aneurysmectomy	K23[45], K24[348]
Myectomy	K24[345678], K373
Primary VAD	K023, K54[189], K56[29]
Pulmonary thrombo-endarterectomy	L04[189]
TEVAR	L2[78 34]
Transplant	K01[1289], K02[124689]
Auxiliary codes	
Valve surgery	K3[01][89], K34[3489], K38[12]
Aortic surgery (abdominal)	L1[89][3456], L2[01][3456], L21[89], L22[23], L253
Congenital	K04[12345689], K05[1289], K06[1234589], K07[12389], K08[123489], K09[12345689], K14[1234589], K17[123456789], K18[123456789], K19[12345689], K20[123489], K24[12], K29[123456789], K305, K315, K33[26], K36[89], K37[1245689], K384, L01[123489], L02[123489], L03[1289], L05[12389], L06[123456789], L07[1234589], L08[12346789], L09[1289], L10[123489], L12[12345689], L23[2347], L69[1234589]
Pacemaker	K60[1567], K61[156789]
Pacemaker—other	K59[123456789], K60[23489], K61[234], K7[234][12389], K724
IABP	K561
Additional—revision	T03[234], K5[45]2, K68[1289], K025
Additional—TAVI	Y494, Y79[12348]
Additional—temporary	Y705
Site	Z32[1234589], Z33[123456789], Z34[123489], Z36[236], Z38[345], Z39[125], Z40[12], Z954, Z985

AF, atrial fibrillation; ASD, atrial septal device; CABG, coronary artery bypass grafting; ECMO, extracorporeal membrane oxygenation; IABP, intra-aortic balloon pump; LV, left ventricle; TAVI, transcatheter aortic valve implantation; TEVAR, thoracic endovascular aortic repair; VAD, ventricular assist defect; VSD, ventricular septal defect.

Table 2 Classification of OPCS codes relative to heart valve surgery

Valve surgery	Aortic	Mitral	Pulmonary	Tricuspid	Unspecified valve
Repair	K26[589], K302, K312	K25[589], K301, K311, K341, K38[123]	K28[589], K304, K314	K27[5689], K303, K313, K342	K30[89], K343
Replacement	K26[1234]	K25[1234]	K28[1234]	K27[1234]	
Other			K346, K362	K345, K361	K31[89], K34[489]

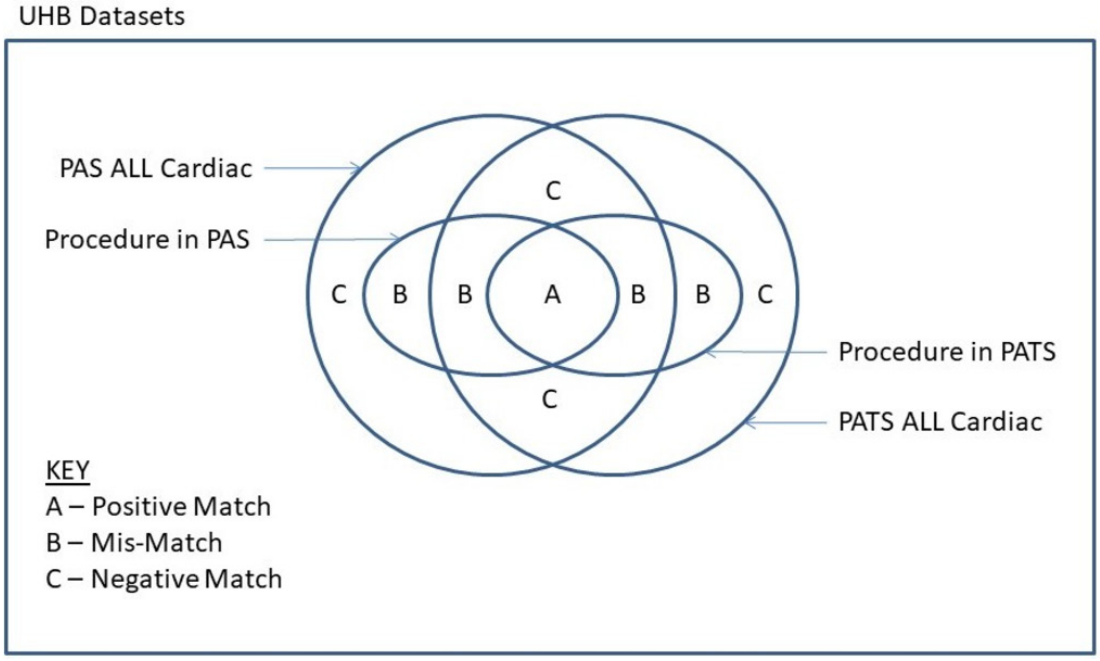

Figure 1 Venn diagram illustrating the matching at local level between PAS (the administrative data set) and PATS (the clinical database). PAS, Patient Administrative System; PATS, Patient Administrative and Tracking System; UHB, University Hospital Birmingham.

prior to the full introduction of PbR in 2009/10 have been addressed via audit.[1 2]

Despite these criticisms, administrative databases are increasingly used in the quality improvement cycle in several countries. In England, administrative data contained in Hospital Episode Statistics (HES) are used quarterly to monitor mortality across trusts by means of the Summary Hospital-level Mortality Indicator, which is the ratio between the actual number of patients who die following hospitalisation and the number that would be expected to die on the basis of average England figures, given the characteristics of the patients treated there.[3]

More recently, the Get It Right First Time (GIRFT) programme has been commissioned by UK's National Health Service (NHS) Improvement with the aim of improving the quality of the service through the reduction of variation between providers. GIRFT used HES among other data sources to benchmark performance. The workflow for cardiothoracic surgery starts from the identification of operations, using Healthcare Resource Groups, a grouping of healthcare services based on the use of a similar amount of resources.[4]

At a more local level, University Hospital Birmingham Quality and Outcome Research Unit has developed a number of Specialty Quality Indicators, reviewing monthly variations in performance across several specialties, with the purpose to enable the quality of care delivery to be measured and improved.[5]

In a paper published recently, we demonstrated that HES data can be used to generate accurate risk prediction models assessing early and late mortality after cardiac surgery.[6] The benefit of using administrative data is that they collect information on powerful prognostic factors which are not usually included in disease-specific registries, such as previous emergency admission to hospital, the presence of medical comorbidities and level of social deprivation. Furthermore, these data are inexpensive, promptly and widely available, and represents real-world treatment settings in unselected populations. Finally, long-term follow-up can easily be obtained through linkage to other central databases.

Table 3 Number of procedures identified in HES in the financial years 2004–2015

	CABG	Valve	Major aortic	Transplant	Other
2004	19605	5530	298	129	722
2005	17713	5529	324	109	737
2006	17602	5819	407	109	801
2007	19217	6394	435	91	873
2008	18208	6770	463	84	917
2009	16177	6517	497	77	1043
2010	15226	6430	595	79	1156
2011	14977	7284	567	92	1109
2012	14119	7320	632	101	1014
2013	14815	7991	664	132	1068
2014	14030	8430	701	110	1139
2015	12829	8658	620	137	1204
Total	194518	82672	6203	1250	11783

CABG, coronary artery bypass grafting; HES, Hospital Episode Statistics.

Table 4 Comparison between PATS (clinical) and PAS (administrative) data at University Hospital Birmingham in the financial years 2012–2015

Procedure group	Procedure	Match negative	Match positive	Mismatch	Total match (%)
CABG	CABG	941	1118	1	100
Valve	Aortic	1340	679	41	94
	TAVI	1923	124	13	91
	Mitral repair	1913	121	26	82
	Mitral replace	1932	104	24	81
	Pulmonary	2010	40	10	80
	Tricuspid	1903	148	9	94
Major aortic	Major aortic	1818	220	22	91
Transplant	Transplant	1993	67	0	100
Other	Other procedures	1676	267	117	70
	AF surgery	1939	85	36	70
	ASD closure	2017	19	24	44
	VSD closure	2047	4	9	31
	LV anneurysectomy	2047	1	12	8
	Myectomy	2049	3	8	27
	Primary VAD	2023	31	6	84
	Epicardial pacemaker vs PAS w/temp flag	1717	4	339	1
	Epicardial pacemaker vs PAS w/no temp flag	2038	4	18	18

Match negative, procedure in neither PATS nor PAS; Match positive, procedure in PATS and PAS; Mismatch, procedure in PATS only or PAS only; Total match, positive/(positive+mismatch).
AF, atrial fibrillation; ASD, atrial septal device; CABG, coronary artery bypass grafting; LV, left ventricle; PAS, Patient Administrative System; PATS, Patient Administrative and Tracking System; TAVI, transcatheter aortic valve implantation; VAD, ventricular assist defect; VSD, ventricular septal defect.

In order to obtain information with a sufficient level of detail, the codes must first be retranslated into a separate catalogue of clinically useful categories. We therefore set out to devise a custom classification of procedures and algorithms based on OPCS, enabling us to identify open heart surgeries from the English administrative database, HES.

METHODS

Custom classification of OPCS codes

Most codes relevant to cardiac surgery belong to OPCS chapters K (Heart) and L (Arteries and Veins), with some additional codes from chapters Y and Z (Methods and Site of Operation, respectively).

We grouped together the codes describing coronary artery bypass grafting (CABG), procedures on heart valves—differentiating between target valve and replacement or repair—major aortic surgery, transplant, ventricular assist devices and other less commonly performed operations. Transcatheter aortic valve implants/replacements (TAVI/TAVR) were identified by a combination of OPCS codes for aortic valve replacement (AVR), and additional Y codes (tables 1 and 2).

Data extraction and validation with local data

Data were extracted from the local administrative database (ie, Patient Administrative System [PAS]) using OPCS, ICD-10 and relevant consultant treatment specialty codes. Clinical data were collected from the local cardiac database (ie, Patient Administrative and Tracking System [PATS], provided by Dendrite) using clinically named procedures. Only admissions containing episodes under a consultant in cardiac surgery were included, while congenital surgery was excluded using both list of OPCS and ICD-10 codes (Q2).

The Jaccard Similarity Index was applied to compare the data selected from these two data sets which included all cardiac patients, and the following subgroups: patient groups as described in the Society of Cardio-Thoracic Surgery (SCTS) bluebook[7] and patients having other named operations with substantive numbers. The Jaccard Similarity Index is the size of the intersection of two sets divided by the size of the union of two data sets and treats each data set in the comparison equally. In this instance

Table 5 Matching in individual trusts, for CABG, AVR, mitral procedures and combinations of these, expressed as ratio of number in HES over number in SCTS bluebook

Trust	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
1	1.16	1.20	0.99	1.41	1.48	1.48	1.08	1.08	1.27	1.23	1.25	1.15
2	1.00	1.05	1.21	1.28	1.33	1.23	1.13	1.00	1.04	1.15	1.21	1.03
3	1.25	1.26	1.22	1.06	1.03	1.04	1.01	0.97	1.01	1.00	0.92	0.95
4	1.01	1.01	1.20	1.27	1.12	1.02	1.02	1.01	1.01	1.00	1.07	0.95
5	1.03	1.05	1.03	1.04	1.03	1.04	1.06	1.05	1.09	1.03	1.05	1.01
6	0.93	0.91	1.06	1.10	1.19	1.23	1.07	1.04	1.02	0.99	0.99	0.96
7	1.00	1.02	1.08	1.10	1.03	1.00	0.99	1.00	1.04	1.15	1.00	1.01
8	1.03	1.02	1.07	1.09	1.12	1.04	1.06	1.01	1.01	1.01	0.98	0.98
9	1.05	1.08	1.08	1.07	1.10	1.06	0.99	1.00	1.02	1.02	0.96	0.96
10	1.05	1.02	1.03	1.04	1.06	1.04	1.03	1.02	1.02	1.01	1.00	0.93
11	1.05	1.07	1.07	1.07	0.98	1.02	1.06	0.99	1.03	1.00	0.97	0.93
12	1.03	0.99	1.05	1.04	1.03	1.02	1.01	1.01	1.02	1.02	1.00	0.97
13	1.02	1.00	1.03	1.01	1.03	1.02	1.03	0.98	0.99	1.02	1.01	1.00
14	1.24	1.06	1.02	1.05	1.00	0.98	1.01	0.98	0.96	0.98	0.94	0.90
15	1.06	1.00	1.01	1.00	1.02	1.01	1.03	1.03	1.01	1.02	0.99	0.95
16	1.03	1.03	1.01	1.02	1.01	1.03	0.98	1.00	0.99	1.03	1.01	0.95
17	1.00	0.98	1.02	1.02	0.99	0.98	1.00	0.98	1.00	1.01	0.97	0.95
18	1.02	1.03	1.02	1.02	1.05	1.05	1.03	1.00	1.01	1.01	1.01	0.60
19	1.01	0.98	0.98	1.02	1.02	0.99	0.99	0.92	0.92	0.97	0.97	0.97
20	1.00	1.02	0.93	0.97	0.94	1.00	0.91	0.93	0.98	1.09	0.99	0.96
21	1.02	1.02	1.01	1.00	1.04	1.01	1.00	0.89	0.94	0.94	0.91	0.93
22	0.95	0.98	1.03	1.02	1.01	0.98	0.99	0.94	0.93	0.91	0.81	0.92
23	0.95	1.00	1.03	1.01	1.01	0.92	0.99	0.90	0.86	1.02	0.83	0.85
24	0.09	0.99	1.03	1.02	1.01	1.02	1.01	0.97	0.98	0.99	0.99	0.99
25		0.04	1.05	1.03	1.19	1.03	1.05	1.02	0.98	0.98	0.98	0.98
26	0.93	0.72	0.67	0.63	0.67	0.68	0.71	0.93	0.96	0.95	0.94	0.61
27			0.02	0.93	0.93	0.95	1.01	0.96	0.99	0.97	0.96	0.95
28				1.03	0.96	0.92	0.98	0.94	0.95	0.94	0.88	0.90
29								0.03	0.98	1.02	0.96	0.90

Trust names have been omitted and assigned a random number. The left column lists all trusts in HES, pseudonymised with a random number.
AVR, aortic valve replacement; CABG, coronary artery bypass grafting; HES, Hospital Episode Statistics; SCTS, Society of Cardio-Thoracic Surgery.

for comparing data selected from the PATS and PAS data sets, the Jaccard Similarity Index was deemed the most natural similarity index to use as the objective was to measure data set overlap (figure 1).

The sensitivity of the data sets alignment was explored under clinical and coding guidance by modifying the set of codes in the PAS data set. The final code selection was validated by a senior cardiac surgeon.

Validation with national clinical data

To check the validity of the selected codes, we cross-referenced national numbers extracted from HES against country-wide figures publicly made available by the SCTS.

Data protection

This study was registered as an Audit with the Clinical Audit Team on the University Hospital Birmingham Clinical Audit and Registration Management System.

For information governance reasons, all figures related to frequencies≤5 were substituted with the symbol *, to avoid potential patient identification.

Patient and public involvement

Patients or public were not involved, as we worked with data already collected by trusts for reimbursement purposes.

Table 6 Matching in individual trusts for CABG, expressed as ratio of number in HES over number in SCTS bluebook

Sum of CABG												
Trust	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
1	1.16	1.20	0.99	1.41	1.48	1.48	1.08	1.08	1.27	1.23	1.25	1.15
2	1.00	1.05	1.21	1.28	1.33	1.23	1.13	1.00	1.04	1.15	1.21	1.03
3	1.25	1.26	1.22	1.06	1.03	1.04	1.01	0.97	1.01	1.00	0.92	0.95
4	1.01	1.01	1.20	1.27	1.12	1.02	1.02	1.01	1.01	1.00	1.07	0.95
5	1.03	1.05	1.03	1.04	1.03	1.04	1.06	1.05	1.09	1.03	1.05	1.01
6	0.93	0.91	1.06	1.10	1.19	1.23	1.07	1.04	1.02	0.99	0.99	0.96
7	1.00	1.02	1.08	1.10	1.03	1.00	0.99	1.00	1.04	1.15	1.00	1.01
8	1.03	1.02	1.07	1.09	1.12	1.04	1.06	1.01	1.01	1.01	0.98	0.98
9	1.05	1.08	1.08	1.07	1.10	1.06	0.99	1.00	1.02	1.02	0.96	0.96
10	1.05	1.02	1.03	1.04	1.06	1.04	1.03	1.02	1.02	1.01	1.00	0.93
11	1.05	1.07	1.07	1.07	0.98	1.02	1.06	0.99	1.03	1.00	0.97	0.93
12	1.03	0.99	1.05	1.04	1.03	1.02	1.01	1.01	1.02	1.02	1.00	0.97
13	1.02	1.00	1.03	1.01	1.03	1.02	1.03	0.98	0.99	1.02	1.01	1.00
14	1.24	1.06	1.02	1.05	1.00	0.98	1.01	0.98	0.96	0.98	0.94	0.90
15	1.06	1.00	1.01	1.00	1.02	1.01	1.03	1.03	1.01	1.02	0.99	0.95
16	1.03	1.03	1.01	1.02	1.01	1.03	0.98	1.00	0.99	1.03	1.01	0.95
17	1.00	0.98	1.02	1.02	0.99	0.98	1.00	0.98	1.00	1.01	0.97	0.95
18	1.02	1.03	1.02	1.02	1.05	1.05	1.03	1.00	1.01	1.01	1.01	0.60
19	1.01	0.98	0.98	1.02	1.02	0.99	0.99	0.92	0.92	0.97	0.97	0.97
20	1.00	1.02	0.93	0.97	0.94	1.00	0.91	0.93	0.98	1.09	0.99	0.96
21	1.02	1.02	1.01	1.00	1.04	1.01	1.00	0.89	0.94	0.94	0.91	0.93
22	0.95	0.98	1.03	1.02	1.01	0.98	0.99	0.94	0.93	0.91	0.81	0.92
23	0.95	1.00	1.03	1.01	1.01	0.92	0.99	0.90	0.86	1.02	0.83	0.85
24	0.09	0.99	1.03	1.02	1.01	1.02	1.01	0.97	0.98	0.99	0.99	0.99
25		0.04	1.05	1.03	1.19	1.03	1.05	1.02	0.98	0.98	0.98	0.98
26	0.93	0.72	0.67	0.63	0.67	0.68	0.71	0.93	0.96	0.95	0.94	0.61
27			0.02	0.93	0.93	0.95	1.01	0.96	0.99	0.97	0.96	0.95
28				1.03	0.96	0.92	0.98	0.94	0.95	0.94	0.88	0.90
29								0.03	0.98	1.02	0.96	0.90

CABG, coronary artery bypass grafting; HES, Hospital Episode Statistics; SCTS, Society of Cardio-Thoracic Surgery.

RESULTS

Patients having cardiac surgery were identified using a combination of procedure codes (having at least one principal code as listed in table 1), age over 18 and consultant treatment specialty included cardiothoracic surgery, cardiac surgery and thoracic surgery. Sensitivity analysis using ICD-10 diagnostic codes in addition to OPCS codes did not improve alignment between PAS and PATS datasets. Applying the custom classification to HES, we identified a total of 296 426 cardiac surgery procedures performed in England over a 11-year period (April 2004–March 2015). Unsurprisingly, the majority were CABG, AVR, mitral valve repair and aortic surgery (table 3).

Operations selected from PAS data set over the more recent period (ie, April 2012—March 2015) largely matched against PATS; the correspondence was especially high for CABG, valve, major aortic surgery, transplant and ventricular assist devices (table 4, figure 1).

In the analysis of coding accuracy at individual trusts level over the years, we report the number of surgeries in the SCTS bluebook divided by counts derived from HES (tables 5–8). In table 5, the ratio of SCTS to HES counts before the full introduction of PbR in 2009 is higher than after 2009. The effect of coding practice is particularly noticeable for trust 1. In 2009, trust 1 had a ratio of 1.48 and in 2010, the ratio fell to 1.08. Other anomalies include the ratio spiking during 2013 for trusts 7 and 20. Overall, results across all years are summarised in table 9. There is a good overall match particularly for CABG (IQR 98.6%–104%), AVR (IQR 105%–118%) and mitral valve replacement (IQR 86.2%–111%).

Table 7 Matching in individual trusts for aortic valve replacement, expressed as ratio of number in HES over number in SCTS bluebook

Sum of AVR												
Hospital	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
2	1.16	1.13	1.56	3.83	4.04	4.70	2.68	1.30	1.32	1.43	1.51	1.20
1	1.42	1.38	1.15	1.43	1.75	1.57	1.12	1.07	1.47	1.50	1.31	1.37
4	1.10	1.27	1.51	1.46	1.22	1.27	1.59	1.32	1.27	1.13	1.18	1.01
9	1.20	1.20	1.21	1.24	1.16	1.22	1.39	1.22	1.07	1.19	1.06	0.87
13	1.12	1.15	1.29	1.19	1.24	1.21	1.18	1.09	1.17	1.16	1.01	0.96
7	1.02	1.12	1.23	1.30	1.10	1.15	1.13	1.08	1.12	1.44	1.06	0.97
3	1.43	1.33	1.43	1.15	1.26	1.24	1.14	1.01	1.01	0.94	0.89	0.85
11	1.10	1.16	1.18	1.09	1.13	1.06	1.22	1.27	1.18	1.15	1.00	1.03
10	1.10	1.07	1.15	1.24	1.11	1.15	1.13	1.10	1.16	1.15	1.07	1.01
16	1.14	1.17	1.10	1.09	1.24	1.21	1.14	1.10	0.95	1.19	1.12	0.93
18	1.14	1.07	1.11	1.11	1.15	1.12	1.18	1.10	1.11	1.08	0.99	1.20
6	0.93	1.04	1.20	1.17	1.07	1.40	1.10	1.19	1.09	1.12	0.98	0.93
5	1.15	1.09	1.20	1.13	1.15	1.11	1.03	1.04	1.09	1.15	1.03	0.96
21	1.14	1.11	1.14	1.22	1.11	1.18	1.14	1.06	0.99	1.05	0.96	0.96
22	1.11	1.16	1.25	1.19	1.12	1.13	1.21	1.06	0.90	0.97	0.84	1.01
23	1.12	1.16	0.98	1.33	1.15	1.06	1.07	1.01	1.16	1.14	0.90	0.88
8	1.18	1.10	1.19	1.08	1.09	1.08	1.13	1.08	1.08	1.06	0.88	0.99
19	1.23	1.10	1.20	1.01	1.03	0.98	1.01	0.94	1.01	1.07	1.09	1.00
12	1.07	1.06	1.09	1.06	1.11	1.17	1.05	1.11	0.99	1.03	1.02	0.87
17	1.28	1.14	1.09	1.05	1.05	1.00	1.09	1.02	1.01	1.08	0.83	0.90
14	1.26	1.18	1.24	1.09	1.03	0.98	0.91	0.99	1.00	1.04	0.89	0.90
15	1.10	1.07	1.10	1.05	1.00	0.99	1.00	0.99	1.05	1.09	0.98	0.83
24	0.11	1.05	1.08	1.10	1.12	1.02	1.04	1.07	1.02	1.02	1.05	1.04
20	1.12	1.10	0.97	0.97	1.00	0.99	0.15	0.07	1.06	1.22	1.02	1.02
25		0.05	1.12	1.05	1.11	1.04	1.14	1.09	1.02	1.14	0.82	0.68
26	1.03	0.77	0.73	0.76	0.73	0.71	0.76	0.92	1.01	0.96	0.83	0.78
27			0.04	1.08	0.99	1.07	1.14	1.13	1.14	1.00	0.95	0.98
28				1.12	1.18	1.01	1.15	0.96	1.04	1.08	0.95	0.93
29								0.02	1.04	1.06	0.81	0.80

AVR, aortic valve replacement; HES, Hospital Episode Statistics; SCTS, Society of Cardio-Thoracic Surgery.

Furthermore, the data extracted from HES using the procedure classification (tables 1 and 2) accurately identified the trusts that are known to provide cardiac surgery in England also including a small number of operations by private providers.

DISCUSSION

There is increasing evidence assessing the use of administrative data in clinical research. Although there are concerns that this data lack reliability, we set out to identify cardiac operations in the major administrative database in use in England and validated this against clinical data. It is our hope that this work will lay the foundations for basis of a risk-prediction model that can be applied to all clinical specialities.

Principal findings

With our classification of procedures from OPCS codes, we were able to identify the most commonly performed adult cardiac surgery operations in an administrative database with over 90% precision. The addition of ICD-10 codes was deemed unhelpful possibly due to only loose alignment between diagnoses and procedures. AVR, for example, may be indicated in patients with valve stenosis, valve regurgitation or endocarditis, while heart failure may be treated by CABG, mitral repair, transplant and more.

Table 8 Matching in individual trusts for mitral procedures, expressed as ratio of number in HES over number in SCTS bluebook

Sum of mitral												
Hospital	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
2	1.11	1.02	1.52	3.00	4.70	2.58	2.09	1.02	0.88	0.89	0.85	0.89
1	1.16	1.09	0.91	1.14	1.10	1.23	0.83	0.95	1.08	1.61	1.38	1.00
8	1.00	1.09	1.13	1.16	1.07	1.09	1.29	1.14	1.16	1.24	1.00	0.94
18	1.15	0.96	1.12	1.50	1.21	0.96	1.04	1.05	1.03	1.18	0.86	1.00
13	1.12	0.94	0.94	1.38	1.18	1.13	1.18	1.26	1.02	1.12	0.87	0.80
21	1.10	1.29	1.24	1.00	1.06	1.32	1.11	1.11	0.94	0.95	0.86	0.65
3	1.37	1.42	1.53	1.05	1.12	1.07	1.14	0.91	0.96	0.81	0.57	0.58
20	1.18	1.05	1.17	1.08	1.00	1.13	0.88	1.16	1.22	0.97	0.91	0.77
7	1.12	1.00	0.96	0.95	1.02	0.87	0.98	1.12	1.16	1.24	0.98	0.87
17	1.19	1.03	1.05	1.47	0.97	1.04	0.91	0.89	1.01	0.90	0.83	0.73
11	1.13	1.12	1.07	0.92	1.09	0.94	0.96	0.95	1.29	0.94	0.83	0.73
16	1.16	1.15	1.21	1.16	1.09	0.93	0.89	0.95	1.03	0.90	0.71	0.69
5	0.89	1.04	1.12	1.10	1.04	0.94	0.92	0.99	0.94	0.94	0.84	0.72
4	0.99	0.93	1.30	1.25	0.83	0.93	1.17	1.07	0.88	0.82	0.68	0.62
12	0.97	0.94	0.90	0.91	0.91	0.98	1.02	0.94	0.85	0.96	0.88	0.78
14	1.11	1.06	0.99	0.80	0.85	0.91	1.01	0.88	0.81	0.87	0.75	0.80
6	0.74	0.85	1.04	1.03	0.88	0.89	0.83	0.81	0.94	0.88	0.94	0.89
23	1.33	0.89	0.88	1.02	1.10	0.82	0.78	0.79	0.92	0.89	0.67	0.63
19	1.03	0.95	0.88	0.78	0.95	0.82	1.00	0.97	0.86	0.90	0.78	0.78
22	0.84	1.13	0.85	1.04	0.84	0.86	0.82	0.84	0.71	0.81	0.80	1.13
25		0.11	1.26	1.16	1.20	1.00	0.84	1.20	1.00	0.95	0.64	0.65
10	1.00	0.92	0.88	0.83	0.82	0.88	1.00	0.65	0.82	0.75	0.71	0.71
9	1.14	1.08	0.83	0.89	0.81	0.74	0.74	0.54	0.68	1.26	0.53	0.69
27			0.12	1.07	0.89	0.91	1.30	0.86	1.18	1.00	0.98	0.91
15	0.84	0.68	0.76	0.82	0.94	0.71	0.56	0.80	0.60	0.69	0.77	0.67
26	0.89	0.81	0.65	0.63	0.73	0.59	0.60	0.80	0.80	0.82	0.74	0.58
28				1.00	0.88	1.00	0.88	0.94	0.83	0.95	0.72	0.79
24	0.17	0.84	0.67	0.89	0.67	0.66	0.66	0.67	0.88	0.72	0.51	0.47
29								0.04	0.86	0.86	0.74	0.76

HES, Hospital Episode Statistics; SCTS, Society of Cardio-Thoracic Surgery.

Strengths and weaknesses of the study

Using OPCS codes only, procedure classification was still high for the most common procedures (ie, CABG, AVR and mitral replacement). Significant over counting in AVR was possibly due to TAVI being classified as regular AVR. A larger degree of imprecision was found when distinguishing between mitral repair and replacement; likely due to miscoding of prosthetic rings as valve implant (table 9).

The main concern when using administrative data sets for research focus is the accuracy of the data, as they are conceived for different purposes. An independent survey of 40 trusts was commissioned by NHS to look into the coding errors that may lead to changes in payment. This survey found an average error rate of 7% (ranging between 1% and 45%), although the financial impact was minimal, and errors were higher in diagnoses than in procedures.[2] The same study identified several critical areas: patient notes, which are often in poor condition; clinicians not differentiating important from less relevant diagnoses, thus leaving the choice to coders lacking medical training; and slow and uneven adoption of new guidelines by coding departments. However, it appears that HES quality has progressively improved in the recent years, likely driven by financial incentives to keep more complete and accurate records after the introduction of the PbR system. This holds even more true for high-cost procedures, which is often the case in cardiac surgery.[8] Moreover, the accuracy of data is ensured via audit of the payment by result

Table 9 Distribution of matching across all trusts in England, divided by procedure groups

	N	CABG	AVR	Mitral	Mitral repair	Mitral replace	CABG+AVR	CABG+mitral	CABG+mitral repair	CABG+mitral replace
1%	0.076	0.074	0.061	0.156	0.000	0.142	0.022	0.000	0.000	0.000
5%	0.882	0.910	0.915	0.666	0.065	0.684	0.797	0.500	0.000	0.556
10%	0.926	0.938	0.986	0.754	0.151	0.843	0.876	0.720	0.075	0.693
25%	0.975	0.986	1.049	0.862	0.478	1.000	0.958	0.829	0.365	0.907
50%	1.000	1.015	1.111	0.957	0.750	1.190	1.016	0.929	0.705	1.106
75%	1.041	1.038	1.179	1.106	0.901	1.467	1.086	1.000	0.875	1.500
90%	1.100	1.093	1.285	1.214	1.044	2.000	1.167	1.121	1.000	2.317
95%	1.148	1.210	1.435	1.301	1.183	2.210	1.228	1.214	1.113	3.075
99%	1.304	1.352	2.960	2.205	2.312	3.333	1.485	1.621	1.416	4.790

AVR, aortic valve replacement; CABG, coronary artery bypass grafting.

system, with standards assessed by the Health and Social Care Information service.[6]

Strengths and weaknesses in relation to other studies

This is not the first study aiming at extrapolating information from administrative databases and comparing them with clinical registries. On the contrary, several reports have been generated in the past decades in various countries, each adopting different techniques to select patients to include.[1 6 9 10] A study from Massachusetts, USA, investigating isolated CABG in a clinical and an administrative database found significant difficulties identifying two corresponding cohorts, because of the way the database and the algorithm was designed (using the ICD-9-CM codes corresponding to CABG lead to the inclusion of cases combined with valvular or other surgery).[9] In other work, Aylin and colleagues analysed mortality after isolated CABG in England, comparing data from the recently established HES and those published by the SCTS. In this case, numbers were not directly compared, only the models derived from each data set were.[1] A similar study conducted in the Netherlands individually identified patients from the clinical registry in the administrative data, using personal information such as date of birth or post code. This led to an only partial matching (77%) due to incompleteness of the data and partial inaccessibility for information governance issues.[10] Finally, last year our group compared a model derived from HES data with EuroScore. The matching was very good (around 98%) although it was only conducted at local hospital level, where data were more readily available and precise.[6]

The contrasting findings among studies can be explained by the inherent differences between each national administrative database, and changes over the years, but also by the study approach. Compared with these previous studies, we achieved a higher level of detail in our classification, with the aim of analysing specific subgroups of operations. In addition, the comparison with the available clinical data, obtained from different sources, provided several layers of control over the quality of our work. It is worth noting that Jaccard similarity could not be used beyond local level, as it would require data linkage by patient rather than linkage of aggregate numbers by hospital and year. Although it was easy to match procedures such as CABG, valve, transplant and VAD, others were difficult to identify and required greater iteration. In context with statistical discrimination procedures, the allocation of OPCS codes to surgical procedures may be regarded as a training phase and hence the degree of matching in tables 5 and 6 is likely to be optimistic, more so for harder to identify procedures.

Meaning of the study

The process we presented here will be replicated for a classification of diagnoses relevant to cardiac surgery, and then applied to the development of a risk model to predict early and late postoperative mortality. Specifically, HES can be linked to the Office for National Statistics mortality data, enabling to track long term, out of hospital follow-up. Possible uses range from risk stratification, to risk prediction, benchmarking and real-time monitoring. Naturally, the better administrative databases become with regard to quality, the more up-to-date, reliable and cheap instruments will be available.

Unanswered questions and future research

An issue which remains largely to be evaluated and hence subject of future study, is the concordance of recorded diagnoses between centres. In the present study, we did not specifically investigate this, although the fact that we were able to correctly identify the trusts providing cardiac surgery services across England is comforting. This system, once appropriately adjusted, can likely be applied to other medical specialties, proving especially useful in those areas where clinical databases are not fully established.

CONCLUSIONS

We described a system to identify the majority of cardiac surgery operations from administrative data, with an improved level of detail compared with the past. In the era of 'big data', administrative data sets make no exception, and given their already widespread use, we believe the best approach is to validate open and transparent standards for their interpretation, addressing the known challenges recently identified by Hand, rather than negating their usefulness with anecdotal or authoritative arguments.[11]

Acknowledgements We would like to thank Mrs V Barnett for data collection and data management, and Mrs C Cottrell and Mr A Price from the coding department for their precious advice on the coding practices and suggestions on the classification.

Contributors This paper is the fruit of a long-standing project, carried by researchers with experience on outcome analysis using clinical and administrative databases. GB, DMN, NF and DP designed the study; GB, DMN, HW, JAM collected the data; all authors participated to the interpretation of the data and revision of the manuscript. GB is the guarantor of the article.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement Data essential for conclusion are included in this manuscript.

REFERENCES

1. Aylin P, Bottle A, Majeed A. Use of administrative data or clinical databases as predictors of risk of death in hospital: comparison of models. *BMJ* 2007;334:1044.
2. Capita. *The quality of clinical coding in the NHS*, 2014.
3. NHS Digital. Summary hospital-level mortality indicator. https://www.digital.nhs.uk/SHMI (Accessed 2Jan 2018).
4. NHS Improvement. Getting it right first time. http://gettingitrightfirsttime.co.uk/ (Accessed 2 Jan 2018).
5. University Hospital Birmingham NHS Foundation Trust. Specialty quality indicators. https://www.uhb.nhs.uk/specialty-indicators.htm (Accessed 2 Jan 2018).
6. Aktuerk D, McNulty D, Ray D, *et al*. National administrative data produces an accurate and stable risk prediction model for short-term and 1-year mortality following cardiac surgery. *Int J Cardiol* 2016;203:196–203.
7. SCTS Bluebook. http://www.bluebook.scts.org/ (Accessed 2 Jan 2018).
8. Edwards Lifesciences. *Cardiac valve & aortic root procedures. 2017/18 national tariff payment system*, 2017.
9. Shahian DM, Silverstein T, Lovett AF, *et al*. Comparison of clinical and administrative data sources for hospital coronary artery bypass graft surgery report cards. *Circulation* 2007;115:1518–27.
10. Siregar S, Pouw ME, Moons KG, *et al*. The Dutch hospital standardised mortality ratio (HSMR) method and cardiac surgery: benchmarking in a national cohort using hospital administration data versus a clinical database. *Heart* 2014;100:702–10.
11. Hand DJ. Statistical challenges of administrative and transaction data. *J R Statist Soc* 2018.